

УДК 004.825

Б.Н. СУДАКОВ, канд. техн. наук, проф., НТУ "ХПИ",
М.В. ИВАНОВА, магистр, НТУ "ХПИ"

МЕТОДИКА РАЗРАБОТКИ ЛИНГВИСТИЧЕСКОГО ОБЕСПЕЧЕНИЯ ЭКСПЕРТНОЙ СИСТЕМЫ

Приведена структура экспертной системы (ЭС), показано место лингвистического обеспечения в составе системы. Предложена структура знаний, которые используются в процессе коммуникативного взаимодействия пользователей и ЭС. Разработана методика создания лингвистического обеспечения ЭС. Ил.: 3. Библиогр.: 10 назв.

Ключевые слова: экспертная система, лингвистическое обеспечение, структура знаний.

Анализ литературы и постановка проблемы. Исследования по автоматизации процессов связаны с бурным развитием вычислительной техники, осознанием того, что ЭВМ, как средства обработки информации, могут сыграть значительную роль в различных областях применения [1].

К настоящему времени накоплен определенный опыт разработки информационных систем, осуществляющих интерпретацию каких-либо данных, рекомендации и т.д. Проведем краткий обзор этих систем.

ВААЛ – предназначена для успешного решения задач маркетинговой коммуникации, когда обычно требуется ответить на ряд контрольных вопросов относительно вербальных элементов коммуникации [2].

Система позволяет: оценивать неосознаваемое эмоциональное воздействие фонетической структуры текстов и отдельных слов на подсознание человека; генерировать слова с заданными фоносемантическими характеристиками; оценивать неосознаваемое эмоциональное воздействие фонетической структуры текстов на подсознание человека; задавать характеристики желаемого воздействия и целенаправленно корректировать тексты по выбранным параметрам в целях достижения необходимого эффекта воздействия; оценивать звукоцветовые характеристики слов и текстов; производить словарный анализ текстов; осуществлять полноценный контент-анализ текста по большому числу специально составленных встроенных категорий и категорий, задаваемых самим пользователем; производить выделение тем, затрагиваемых в текстах, и осуществлять на основе этого автоматическую категоризацию; производить эмоционально-лексический анализ текстов; настраиваться на различные социальные и

© Б.Н. Судаков, М.В. Иванова, 2013

профессиональные группы людей, которые могут быть выделены по используемой ими лексике; производить вторичный анализ данных путем их визуализации, факторного и корреляционного анализа [3]. Схема представления знаний основана на фреймах, а механизм вывода – на последовательном порождении и проверках. Система реализована в виде набора DLL-библиотек, которые подключаются к наиболее популярному текстовому процессору Word for Windows.

Система поддержки принятия диагностических решений – СППДР предложена для диагностики комы, ишемической болезни сердца, острого живота. Содержит алгоритмы экспертной классификации для отсеивания неинформативных признаков [4]. Имитирует выбор человеком наиболее информативных признаков и их комбинаций. Задача системы, прежде всего, состоит в том, чтобы ставить пользователю наиболее информативные вопросы и накапливать знания по целым группам возможных ситуаций. Каждый из признаков имеет несколько градаций (в среднем – 4), которые устанавливает и ранжирует эксперт [5]. Разработанный алгоритм классификации признаков позволяет построить базу знаний (около 2000 возможных ситуаций) за 12 ч. работы эксперта с системой, при этом эксперту приходится классифицировать только около 400 различных ситуаций, так как программа сокращает перебор.

DRUG INTERACTION CRITIC – помогает врачам решать, как назначать лекарства на фоне других лекарств. Система определяет как нежелательные, так и благоприятные взаимодействия, объясняет, почему взаимодействия происходят, отвечает на вопросы о взаимодействиях и предлагает корректирующие действия для устранения побочных эффектов. Является БЗ о типах лекарств. Организована в виде иерархии фреймов, содержащих информацию, связанную с конкретными лекарствами, например об их химическом средстве, избирательном накоплении в конкретных органах и тканях и характеристиках взаимодействия. Знания о механизмах взаимодействия организованы посредством фреймов в соответствии с одним из четырех типов механизмов: физико-химическим, фармакодинамическим, фармакокинетическим и физиологическим. Система содержит ограниченный интерфейс на ЕЯ, включающий орфографический корректор. Реализована на языке Пролог в Политехническом институте штата Виргиния на уровне исследовательского прототипа.

IBM Watson – компьютерная система IBM Watson основана на технологии DeepQA и предназначена для извлечения и анализа информации из текстов на естественном языке, а также для быстрого поиска ответов на вопросы. DeepQA основана на статистическом подходе

в компьютерной лингвистике, то есть для построения когнитивных систем используются мощности вычислительной техники. Схема управления системы обеспечивает некую разновидность прямой и обратной цепочек рассуждений, реализованных применением плана действий, причем каждое предполагаемое утверждение помещается в план в соответствии с приписанным ему приоритетом [6].

Проведенный обзор и опыт эксплуатации существующих информационных систем позволяет сделать следующие выводы:

1. Автоматизация различных процессов может быть реализована в двух направлениях: первое – связано с автоматизацией обработки данных. Второе – призвано оказать помощь в осмысливании большого потока информации о чем-либо за счет выделения полезных признаков, накопления огромного опыта экспертов с наглядным отражением результатов анализа.

2. Задачи относятся к слабоструктурированным задачам, решение которых осуществляется в условиях неопределенности. Знания, закладываемые в систему, отражают субъективное восприятие экспертом и являются в основном эвристическими, неполными и противоречивыми [7].

3. "Старение" моделей, заложенных в систему, требует постоянного их совершенствования вследствие возникновения новых методик, препаратов и т.д. Это приводит к необходимости проведения постоянных доработок программного обеспечения, что является весьма проблематичным в системах, алгоритмы которых построены по "жесткому" принципу.

4. Отсутствует единый подход к формализации знаний о предметной области. Каждый разработчик использует те или иные формализмы с учетом специфики решаемых задач.

5. Ограничены возможности взаимодействия с разработанными системами непрограммирующих пользователей на всех этапах.

6. Общение с ЭС осуществляют различные типы пользователей (оператор, эксперт, когнитолог, администратор). Существующие методики разработки ЛО направлены на создание собственных языковых средств для каждого типа пользователей, а также индивидуальных словарей. Такой подход к разработке ЛО приводит к отсутствию системности, значительному дублированию языковых единиц текста, а также словарной информации, затруднению общения пользователей между собой, ведет к неоправданному расходованию ресурсов вычислительных средств на реализацию языка.

Это объясняется следующими причинами:

– применение неестественных для человека языков общения;

– сложность и трудоемкость подготовки исходных данных, а отсюда и длительность решения задач;

– отсутствие в ряде случаев обратной связи от ЭВМ к пользователю, что не позволяет контролировать состояние системы и порядок решения задач;

– изолированность задач приводит к тому, что сценарии диалога и их языковое содержание повторяются во многих задачах, что увеличивает требуемый объем памяти ЭВМ [8].

С учетом изложенного, можно заметить, что в рамках традиционного подхода к созданию информационных систем намечился некоторый тупик в разработке математического, программного и лингвистического обеспечения этих систем. Во многом это связано с общей структурой знаний, которые используются в процессе коммуникативного взаимодействия пользователей и ЭС [9]. По мнению авторов можно выделить пять уровней знаний:

Зн1 – знания пользователей;

Зн2 – материализованные знания (руководства, инструкции и т.д.);

Зн3 – поле знаний (полуформализованное описание Зн1 и Зн2 в виде рисунков, таблиц, схем, сетей, диаграмм, описаний на ограниченном естественном языке);

Зн4 – формализованные знания на языках представления знаний;

Зн5 – база знаний на машинных носителях информации.

Данную концепцию иллюстрирует рис. 1.

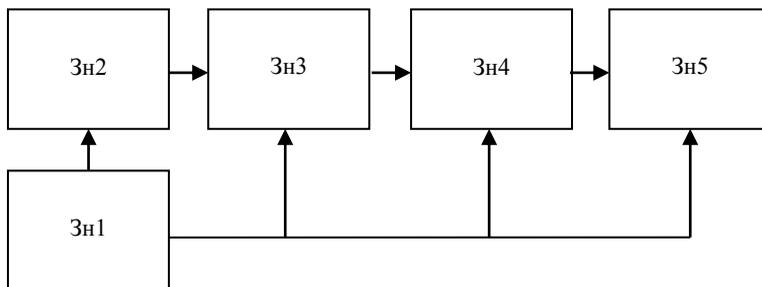


Рис. 1. Общая структура знаний

Совокупность Зн1 и Зн2 образуют знания о предметной области. В ЭС используется то, что удалось извлечь и формализовать из Зн1 и Зн2. Для взаимодействия пользователей с ЭС важным оказывается разработка поля знаний, когда создается полуформализованное описание решаемых задач, основных понятий предметной области и связей между ними. В

дальнейшем структура знаний представляется в виде формализованного описания Зн4.

На каждом из перечисленных уровней знаний используются свои языки. На уровнях Зн1 и Зн2 используется, как правило, профессиональный естественный язык – единственный метаязык, позволяющий описать практически всю предметную область. Уровень знаний Зн4 описывается с использованием некоторого языка представления знаний. В качестве языка на уровне Зн5 выступает внутренний язык (ВЯ) компьютера.

Цель статьи. Разработать методику, направленную на разработку единого языка общения пользователя системы, ограниченного синтаксически, семантически и прагматически рамками решаемой задачи.

Основные положения этой методики приведены на рис. 2 и сводятся к следующему.

На основании задач, решаемых пользователями при взаимодействии с системой, а также требований к лингвистическому обеспечению обосновывается выбор языков взаимодействия, которые необходимо иметь в системе [10]. Проведенный анализ показал, что наиболее целесообразно иметь в системе один язык взаимодействия, состоящий из двух подмножеств: подмножества ограниченного естественного языка (ОЕЯ) и подмножество ВЯ. Преобразование одного подмножества в другое осуществляет транслятор, который в дальнейшем будем именовать лингвистическим процессором (ЛП).

Исходя из требований и принципов построения ЛО, разрабатывается общий подход к структуризации знаний о предметной области (Э5) и производится семантическая классификация понятий ПО (Э6), на базе которых формируется словарь (Э7).

Используя выбранный подход к представлению знаний, синтезируются языковые структуры (Э8), позволяющие адекватно описывать ПО.

Для формального описания конкретных языковых выражений выбирается математический аппарат (формальные грамматики и их модификации, деревья синтаксического подчинения и др.), позволяющий построить модель языка (Э9).

После разработки модели языка (Э10) с использованием выбранного аппарата оценивается качество лингвистического обеспечения (Э11), сложность алгоритмов, а также соответствие ЛО предъявленным требованиям. Если оценка качества не удовлетворяет предъявленным требованиям, то уточняется математический аппарат для описания языка

либо изменяются требования к ЛО, что позволит создать систему взаимодействия для ЭС.

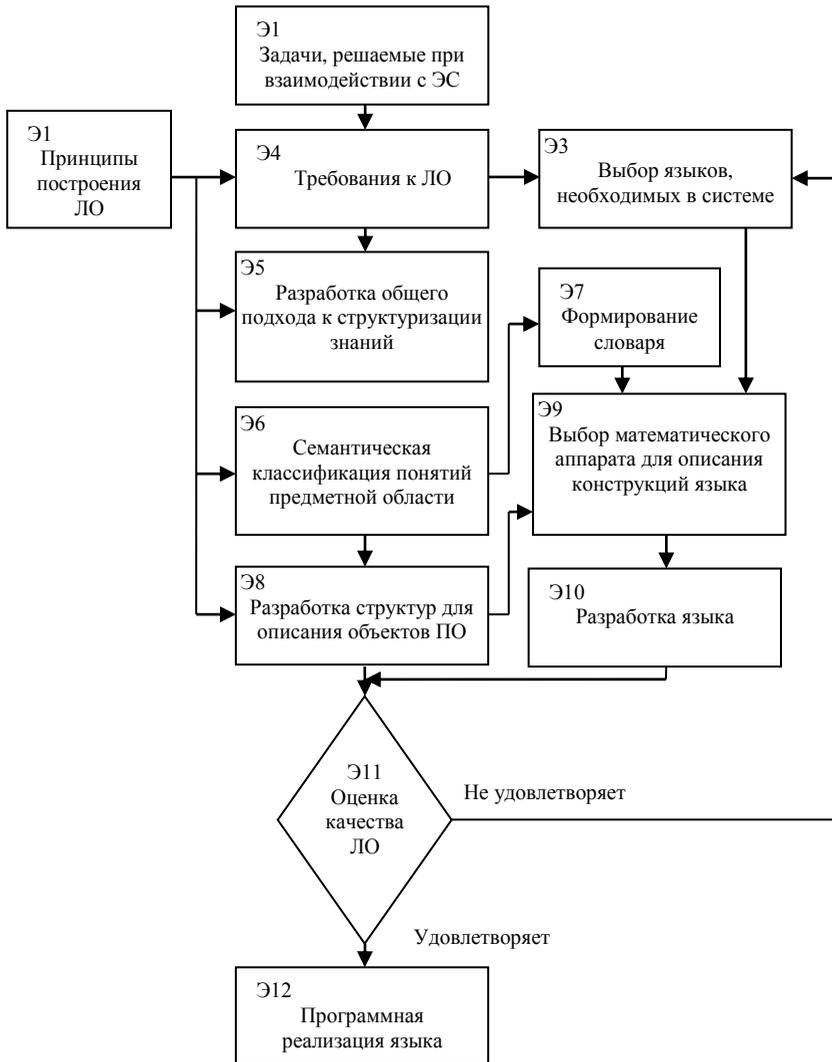


Рис. 2. Методика разработки лингвистического обеспечения

Завершающим этапом разработки ЛО является программная реализация предложенного языка взаимодействия (Э12).

Вывод. Предложенная методика разработки лингвистического обеспечения, позволяет осуществить выбор необходимого языка в системе, математически точно описать язык взаимодействия и адекватно отразить формально-логические основы моделей знаний о предметной области.

Список литературы: 1. ДСТУ 2481-94. Системи оброблення інформації. Інтелектуальні інформаційні технології. Терміни та визначення. – К.: Держстандарт України, 1994. – 30 с. 2. *Башмаков А.И.* Интеллектуальные информационные технологии: Учеб. пособие / *А.И. Башмаков, И.А. Башмаков.* – М.: Изд-во МГТУ им. Н.Э. Баумана, 2005. – 398 с. 3. *Гуцин А.Н.* Экспертные системы: учебное пособие. Балт. гос. техн. ун-т / *А.Н. Гуцин, И.А. Радченко.* – СПб., 2007. – 154 с. 4. *Джозеф Джарратано* Экспертные системы: принципы разработки и программирование / *Джарратано Джозеф, Гари Райли.* – М.: Издательский дом "Вильямс", 2006. – 1152 с. 5. *Жданов А.А.* Автономный искусственный интеллект / *А.А. Жданов.* – М.: Бином, 2009. – 154 с. 6. *Люгер Дж.* Искусственный интеллект: стратегии и методы решения сложных проблем: пер. с англ. / *Дж. Люгер.* – М.: Вильямс, 2003. – 864 с. 7. *Стюарт Рассел, Питер Норвиг* Искусственный интеллект: современный подход (AIMA) / *Стюарт Рассел, Питер Норвиг.* – М.: Издательский дом "Вильямс", 2006. – 1424 с. 8. *Субботин С.О.* Подання й обробка знань у системах штучного інтелекту та підтримки прийняття рішень / *С.О. Субботин* // Навчальний посібник. – Запоріжжя: ЗНТУ, 2008. – 341 с. 9. *Шамис А.Л.* Поведение, восприятие, мышление: проблемы создания искусственного интеллекта / *А.Л. Шамис.* – М.: Едиториал УРСС, 2005. – 132 с. 10. *Freeman J.* Argument Structure. Lawrence Erlbaum Associates / *J. Freeman.* – 2005. – 215 с.

Поступила в редакцию 22.06.2013

После доработки 13.11.2013

*Статью представил д-р техн. наук, проф. НТУ "ХПИ" Серков А.А.
УДК 004.825*

Методика розробки лінгвістичного забезпечення експертної системи / Судаков Б.М., Іванова М.В. // Вісник НТУ "ХПІ". Серія: Інформатика і моделювання. – Харків: НТУ "ХПІ". – 2013. – № 19 (992). – С. 153 – 159.

Наведено структуру експертної системи (ЕС), показано місце лінгвістичного забезпечення у складі системи. Запропоновано структуру знань, які використовуються в процесі комунікативної взаємодії користувачів і ЕС. Розроблено методику створення лінгвістичного забезпечення ЕС. Лл.: 2. Бібліогр.: 10 назв.

Ключові слова: експертна система, лінгвістичне забезпечення, структура знань.

UDC 004.825

Methods of development of linguistic expert system / Sudakov B.N., Ivanova M.V. // Herald of the National Technical University "KhPI". Subject issue: Informatics and Modeling. - Kharkov: NTU "KPI". – 2013. – № 19 (992). – P. 153 – 159.

The structure of the expert system (ES), the place of the linguistic support in the system. The structure of knowledge used in the process of communicative user interaction and ES. The method of creating linguistic support ES. Figs.: 2. Refs.: 10 titles.

Keywords: expert system, linguistic support, structure of knowledge.